INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11 N12967
July 2012, Stockholm, SE

Source 3DG
Status Working draft 2.0
Title Augmented reality reference model
Editors Jae Joon Han, Christine Perey, Marius Preda

Contents

Page

Introduction

Editor’s note: Introduction would provide motivations of augmented reality, the purpose of this part; and the list of contributors.

Digital information is ubiquitous. It refers to anything of interest to those who create content (authors) and their audiences, the consumers of information. Some content and subjects of interest refer to and/or are associated with people, places and objects in the physical world. Content that refers to or is associated with people, places and objects in proximity to a user can be of particularly high value for decision-making, learning or entertainment.

Augmented Reality is a term that describes a suite of technologies and systems permitting users to “consume,” via their sense of vision, auditory perception or other means (e.g., haptic), the data (aka digital information stored as bits), in close association with their environments. With Augmented Reality, the digital data is synchronized with people, places or objects on which the user has elected to focus full or partial attention.

The use of one or more sensors to detect the user’s context and the focus of a user’s full or partial attention are key elements in Augmented Reality when the result of that detection and subsequent processes can be described as an “annotation” of the physical world with digital data.

People can benefit from Augmented Reality-enhanced “experiences” to achieve an unlimited number and types of objectives. Due to the infinite range of potential content and targets, users and objectives, a framework for communication and to establish common understanding is necessary.

The purposes of the Augmented Reality Reference Model include:

- Basis for coordinating activities and understanding of the AR community and technology
- Provide insight into current state of AR
- Communications between acquirers and developers of technology
- Input to component design and development
- Resource for defining application specific architectures.
- Allow coordinated development of standards relevant to AR

The Augmented Reality Reference Model

- defines an authoritative basis that outlines the set of principles, terms and their precise definitions, a generic model of mixed/augmented reality system and its components and interfaces.
- provides the basis for deriving the contents model, its proper abstraction level and required components.

The Augmented Reality Reference Model is

- Independent from algorithms such as
  - Recognition/tracking, Rendering, …
- Independent from sensors used such as
  - Camera, RFID, Kinect, Marker based, Image patch based, …
- Independent from terminals and infrastructure used such as

– Desktop, mobile,
– Projection, HDM, holography

Several standards are available for defining a Reference Model. These include IEEE 1471, DODAF, TOGAF and Zachman. The present work is based on the ISO/IEC 10746-1 standards for Open Distributed Processing (RM-ODP). The model specifically calls for separation of concerns into 5 viewpoints: enterprise, information, computational, engineering, technology. The following sections describe in details four of the five views, the "technology" view being out of the scope of this standard

1 Scope

The Augmented Reality Reference Model (ARRM) identifies and describes the “concerns” from four separate views: Enterprise, Information, Computational and Engineering. The user’s position and point of view (orientation in space) are not fixed in time, although at any point in time the position relative to real world is known and can be constant (unchanging).

The ARRM scope concentrates on audio/visual augmentation of the user’s live "hear"/“view” of the real world with digital assets. The perception of the physical world can be “direct” or “indirect.” Indirect live view is that which is transmitted to the user from a remote location.

marius preda 7/6/12 10:23 AM
Comment [1]: Check the term with an EN native

2 References

Editor’s note: the references will be added in a future edition.

3 Augmented Reality Reference Model

Editor’s note: The system architecture of the reference model is added, the diagram needs to obtain more visibility.

3.1 *Reference Model Concepts*

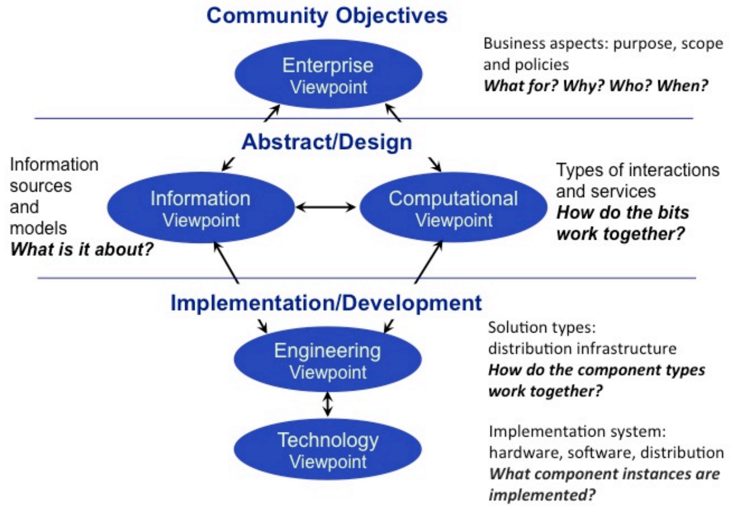


Figure 1 – RM-ODP Viewpoints

Editor's note: Figure should be restyled.

3.2 *AR RM Introduction*

An Augmented Reality System allows real-time access to multimedia information like text, video and audio, synchronized with the end-user context.
A typical Augmented Reality System may consist of elements illustrated in Figure xxx.

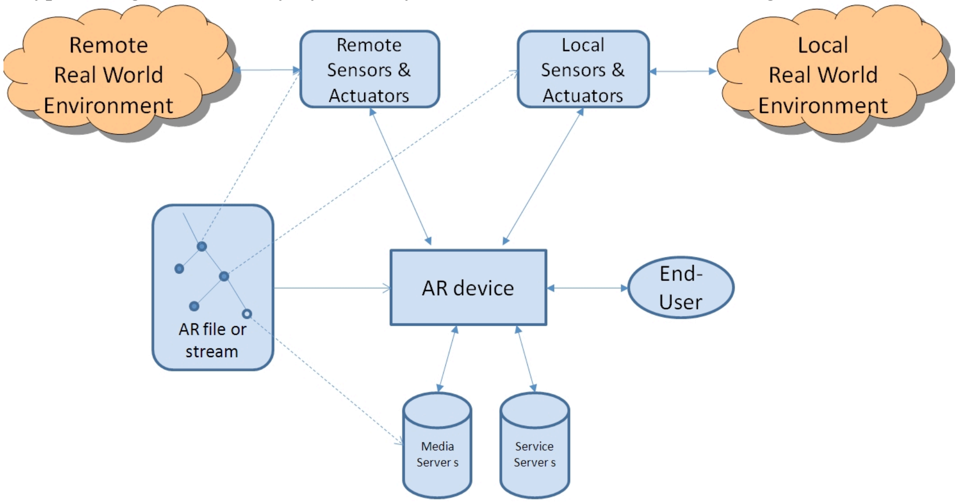


Figure xxx. Possible configuration of an Augmented Reality System. (Marius to do)

Table 1 identifies the specific ARRM topics that are described in the four viewpoints.

Table 1 – Viewpoints in the ARRM

Viewpoint	Viewpoint Definition	Topics in ARRM
Enterprise	• Articulates a "business model" that should be understandable by all stakeholders • Focuses on purpose, scope, and policies • Community Objectives • Business aspects: purpose, scope and policies	• Actors • Components defined for enterprise objectives • Desirable Characteristics (top-level requirements)
Information	• Semantics of information and information processing • Describes the structure and content types of supporting data • Information sources and models • What is AR content about?	• Spatial Registration • Information Content of Virtual Objects • Context Dependent Viewing
Computational	• Identify interfaces that allow for distribution • Types of services and protocols • How do the bits work together and apart?	• Use cases • Identify services and interfaces
Engineering	• Identification of component types • To support distributed interaction between the components • How do the components work together?	• Engineering Components • Interfaces between components • Device Internal Interfaces • Encoding Formats

AR RM Foci

- Enterprise view – Business roles
  - Identify major actors and roles
  - User Scenarios
- Information view – Registration
  - Coordinate Reference Systems
  - Geometry, Symbols, Feature types
- Computational view – Behavior
  - Use cases
- Engineering view – Software
  - Device stack/APIs
  - External networks: Internet, NFC

In order to remain independent of all technological implementations and approaches, the current AR RM is limited to the four viewpoints described in Table 1 above.

Christine Perey 7/6/12 10:23 AM
Comment [2]: Purpose scope and policies of WHAT? The enterprise? This is repeated below in the same cell of the table: Business aspects: purpose, scope and policies but it is not more clear to me

Christine Perey 7/6/12 10:23 AM
Comment [3]: Which community? Enterprise?

Christine Perey 7/6/12 10:23 AM
Comment [5]: I rcommend we use the word "digital" rather than "virtual" but really I think you mean here not "objects" but the augmentations, according to the term as defined during June 16 meeting Session 12 of Taiwan meeting

Christine Perey 7/6/12 10:23 AM
Comment [4]: Not sure if this is the question…the answer is the reality but otherwise?

Christine Perey 7/6/12 10:23 AM
Comment [6]: Distribution of computational tasks? Or distribution of data?

Christine Perey 7/6/12 10:23 AM
Comment [7]: This feels like engineering but anyway, could be described in the process figure developed June 16 meeting Session 12

marius preda 7/6/12 11:36 AM
Comment [8]: This is a repetition of the table above. We should remove it and complete the table

The content of the AR RM is developed by the ISO/IEC JTC 1 SC29 WG 11 in collaboration with experts and contributors in several domains and Standards Development Organizations. [We can use language here from the position paper]

–

marius preda 7/6/12 10:54 AM

Comment [9]: I think that this text is not appropriate in the document; the readers of the document care less on how we internally organized to create it

3.3 *Enterprise Viewpoint for AR-RM*

Editor's note: need to provide the scope of enterprise viewpoint.

NOTES:

- Viewpoint definition
  - Articulates a "business model" that should be understandable by all stakeholders;
  - Focuses on purpose, scope, and policies.
  - Community Objectives
  - Business aspects: purpose, scope and policies
- Topics in Viewpoint for AR
  - General Definition of AR
  - Components defined for enterprise objectives
  - Desirable Characteristics (top-level requirements)

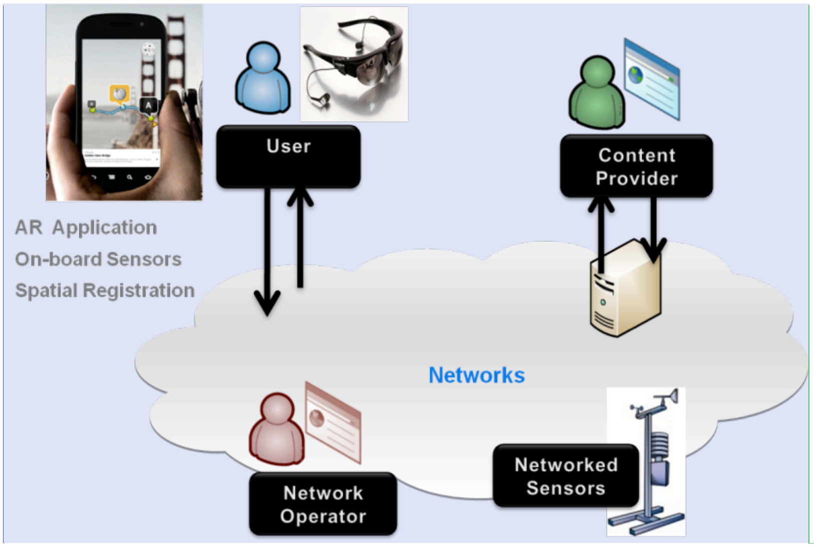
END NOTES

The enterprise specification describes the objectives, policies and requirements of the concerned Augmented Reality System. The requirements and policies of the service are derived from the parties involved. They can be classified according to their role:

- Content creator: a designer (person or organization) that creates multimedia content (scenes, objects, …) implements the application behavior,
- Telecommunication operator: an organization that commercially manages the communication of information between the other actors
- Device manufacturer: an organization that created the device used as end-user terminal, in charge of realizing the augmentation
- Service provider: an organization that commercially manages the services
- End-user: a person who uses services in order to satisfy information access and communication needs

3.3.1 Business model of Augmented Reality

By means of their digital information display and interaction devices, such as smart phones, desktops and tablets, users of Augmented Reality hear, see or feel digital information associated with natural features of the real world and with their context, in real time.



3.3.2 Goals of Augmented Reality

In the table below, the requirements for the success of AR are captured. There are requirements from the user perspective, the system perspective and the developer's perspective in this table.

Create one or more tables.

- Real time contents registration:
  - register digital object (content that is relevant) with real world, based on context (using any means such as but not limited to geospatial coordinates, vision, other…)
- Multiple environments:
  - built environment or natural environment,
  - indoor/outdoor
- Multiple scales
  - Precise augmentation (mm, for repair, heathcare/medicine, etc)
  - At a distance (identification of buildings, cities or mountains)
- Visibility
  - Human observable targets
  - Invisible (infrastructure under the road, in the building, inside a motor)
- Modularized technology
  - To encourage innovation across and from multiple suppliers
  - allows the technology and/or the user experience to be incorporated into existing applications.
- Enable mass content creation by non-programmers
  - Content from anywhere
  - Like the web

Christine Perey 7/6/12 10:23 AM
Comment [10]: Needs to be reworked. No reference to an AR application, for example.

sait 7/6/12 10:23 AM
Comment [11]: Business Entities should be defined their terms and their scopes.

Christine Perey 7/6/12 12:50 PM
Comment [12]: There should be an introduction that specifies if the requirements of AR enumerated here should be only from the enterprise perspective, or if the developer of AR experiences is also an enterprise.

Christine Perey 7/6/12 12:50 PM
Comment [13]: Is "real time content registration" different from "real time recognition" and then (in a separate step) "tracking" of the recognized content? Might be important to separate these two.

Christine Perey 7/6/12 12:50 PM
Comment [14]: Check the definitions. Link terms to these definitions

3.3.3 Purpose [AR scenarios in user language (Enterprise User Scenarios)]

What is a User Scenario Category?

A category of use cases is composed of a “core use case” and many possible permutations. The category description includes the most simplified use case and short descriptions that show how the “core use case” can be adapted to multiple vertical markets and may be extended to support new features.

All Use Cases have several attributes in common

In the description of the AR use case, there does not need to be a detailed (or any) specification of the user’s device. The reason is that the standard that will result from the use case must be sufficiently broad to be used from any device with the sensors and processing ability to complete the tasks requested of it by the user. In other words, it doesn’t matter if the user has a tablet, a smartphone, a heads-up display with connection to a local processing unit. With the exception of standards that concern themselves directly with the connectivity and communications between the devices and information “sources,” the Internet where the processing is done in the cloud, or any other scenario for connectivity, there should not be any reference to where and how much data is stored for the completion of the use case.

3.3.3.1 The Guide Use Case Category

The simplest and most fundamental use case is one in which the real world provides a user interface for a person who is asking simple questions, most often a series of questions that lead to the successful learning, completion of a task or arrival at a destination.

The emphasis of a “guide” is on a point of interest and usually a process (step by step sequence). In the Guide, the user turns on the device, points it at a target (or in a direction) and queries the system to select the question that the user wants to answer or address.

3.3.3.1.1 *Sample Guide Use Case-Geo-position*

The user is of any age, gender or cultural background. The user is new to a particular location. In the scenario the purpose of the user’s arriving at the site is to explore cultural heritage. In advance, using the latest technologies, the site’s management has created interactive experiences for mobile AR platforms using historical information in one or more of these formats: text, sounds (speech), images, 3D models and animations.

The user’s geospatial position and the user’s orientation are determined (method is not relevant). Given the position and orientation, the system welcomes the user to the site. The system may prompt the user to determine some user preferences, but this is not desirable. The system can prompt the user to enter demographic data (age, gender, origin) that can be used for selection of preferred tour. In addition, based on preferences, the system can operate only with audio, with audio+video, with 3D models or only with text labels (augmentations). The system only proposes to the users those options that are supported by the user’s device.

The system also prompts the user to determine the level of interest in different domains (e.g., agriculture, architecture, lifestyle, commerce, entertainment, etc). This information can be organized in small chunks (atomic size?) for the user to explore to the level that they wish. Here the user could use the "more" or "next" button.

Once configured, the system proposes a starting point on the site. Assumption here is that there are unique "tours" that each has its own starting point. The user is guided to the starting point using arrows on the screen or audio instructions.

The user’s arrival at the starting point is detected. The user is prompted to start the tour. A context-sensitive menu should be available as the user move near points of interest, to prompt the user. The user then follows directions on the screen (defined by a story teller/architect of the tour) and at any point can interact with information assets provided as part of the experience (see above regarding the "next" and "more" note above).

By pointing at the building (or where a building was), a reconstruction of the original site (or the building) appears. A photo overlay is one option to bring up the visuals. If the user has selected that they wish to see 3D models, they will see 3D. The 3D model remains registered to the site as the user position changes (tracking the building staying in view relative to the original building as long as the orientation indicates the user's attention is on the same point of interest).

The distance of the user from the object of interest needs to be considered in how to design the interaction between the user and the data. Expressing the content as compact representations is useful here. This shows different types of information aggregation for AR authors.

The user’s orientation changes. Now it is focused on a monument (e.g. statue). This causes a label to appear with information about the date of the monument’s erection, the artist who made it and there is a short animation showing the artist at work (re-enactment of history).

The user wants to be in a photo with the statue. This is done using the user-facing camera. The composition results with the statue, the user and the fictional rendering of the artist leaning on the other side of the statue. The user sends this photo montage to friends via a social network. This portion of the use case can be extended to include capturing a video of the user, location and digital element.

Upon conclusion of the exploration of the site (but this can be available at any time), the user is prompted to produce a comment or to rate a particular component. The user’s rating becomes an augmentation that can then be viewed at another point (e.g., by the curator of the cultural heritage site). The important point is that the user can create and attach an augmentation.

3.3.3.1.2 *Sample Guide Use Case-Computer Vision*

The user can be of any age, any gender or cultural background, in any location at any time of day or year. The focus of the actor/user’s attention is a medium to large piece of machinery (a coffee maker, a washing machine for laundry, a office photocopier, a modern engine) that has continuous power supply, a panel or area on which there are LED indicators of status, an internal computing system and many complex moving parts. The user does not have a manual for configuration, use or repair. The machine’s manufacturer has developed an AR manual for

diagnostic and repair. The AR-assisted manual has 3D and planar recognition algorithms and contains step-by-step instructions.

The device is not performing the task it should be and requires user intervention as indicated by a red light on the panel that is blinking at 1 second intervals. The user orients the device that has a front facing camera, computational resources, communications capabilities and a touch sensitive display, at the machine.

The user communicates with software on the device or in the cloud that they wish to search for an interactive, digital user manual about this machine. The user case does not describe whether the menu is multi-touch, speech controlled or any other details concerning the user interaction for starting, and communicating desires prior to the use of AR.
The user receives an indication from its device that it is searching for a manual but has not determined which manual since it does not know the model of machine the user wishes to receive/use. The software prompts the user to approach the object of interest.

When pointed at the machine, the device's camera sends video of unknown resolution at an undetermined frame rate to a processing system. The system analyzes the video stream, recognizes the point of interest and highlights it with a visual overlay. It can prompt the user by way of speech output, displaying messages on the screen (text or arrows) that the user must change positions with respect to the machine.

When the machine is recognized, the manual is prepared for viewing/interaction (could be local or in the cloud) and the user receives a signal that the first step of diagnostics is ready to begin. The user gives the device a signal (touch, speech, or a gesture) to begin.

A series of diagnostic tests are given to the user to perform. [The use case description can go into more detail here.] Each step is communicated by way of an on-screen (or optional auditory) indicator of successful completion of a step. The next step is then shown and the instructions given. The user gives a signal to the system by way of touching the screen, a gesture or an auditory input (speech).

At the end, the light (LED) turns green and is no longer blinking.

3.3.3.2 Create Use Case Category

Editor's note: provides the specific sample use cases similar to guide use case.

The Create use case category involves the real world, a user's actions with a device. The user's objective in the Create use case is to express an opinion, provide additional thoughts or questions, in other words to "author" something in the form of text, image, video or audio recording and to attach this original (user generated) information to a real world target.

In the Create use case category, the real world target (the Point of Interest of the user) can be anything that the device can sense repeatedly. An example of a Point of Interest that cannot be, in this category, considered a real world target is something ephemeral, such as a wave on the surface of water or a cloud.

In the Create use case category, all conditions with respect to viewing ("consuming") the annotation are equally included. In other words, defining who can view the augmentation (only

sait 7/6/12 10:23 AM
Comment [15]: if this is to share information with others, what is the difference between the play use cases?

the author, only those designated by the author, anyone/public) can be part of a use case within the use case category.
Specific use cases within the broader Create use case category can also embellish aspects of security, curation/review, and distribution of the content.
A use case within the Create use case category case specify where the platform for acknowledging the user's actions is based (local, in the cloud, or both) and the parameters related to the transmission, storage and processing of the user-generated augmentation.

3.3.3.3 Interactive Multi-users Case Category

Editor's note: provides the specific sample use cases similar to guide use case. It seems to be too ambiguous to categorize.

The Play use case category encompasses all use cases in which there is the real world, devices, digital content (augmentations) and two or more users interacting with one another in real time. In the Play use case category, there is no apriori limit to where the users are in the world with respect to one another.

A specific Play use case can specify the distance between the users (proximity) in meters. Other Play user cases can specify the categories of objects that constitute the focus of attention. For example, there are use cases in Play category involving manufacturing, repair, maintenance of machinery, infrastructure or some stationary, man-made object.

Play use cases are not limited to games. There are use cases in the Play category that can describe a military scenario in which two or more people target a common person, place or object and then see what the other sees.

sait 7/6/12 10:23 AM
Comment [16]: I think that the multi-user case should have its own difference other than create use case. I am not sure if I understand it correctly or not.

3.4 *Information Viewpoint for AR-RM*

The Information viewpoint defines the semantics of information and information processing. By the "semantics of information" we are referring to the categorization (separation into discreet categories) of information types. The information viewpoint also defines the information processing. This refers to the stages of possible "treatment" of the information, such as the extraction of natural features from the data that is streaming from the camera, the microphone, the IMUs or another source of real time observations.

The information viewpoint also describes the structure and types of data found in Augmented Reality. It defines terminology for the information sources (sensors, users interaction such as touch or movements?) and models that the other viewpoints use for computation.

In addition, the information viewpoint facilitates the communication of concepts about content used in AR experiences, content that is the target (focus of user attention and the context), the reference which is the extraction of all unique characteristics of the target, and the augmentation (digital object that is anchored to the target).
NOTES:

- Viewpoint definition
  - Semantics of information and information processing
  - Describes the structure and content types of supporting data.
  - Information sources and models

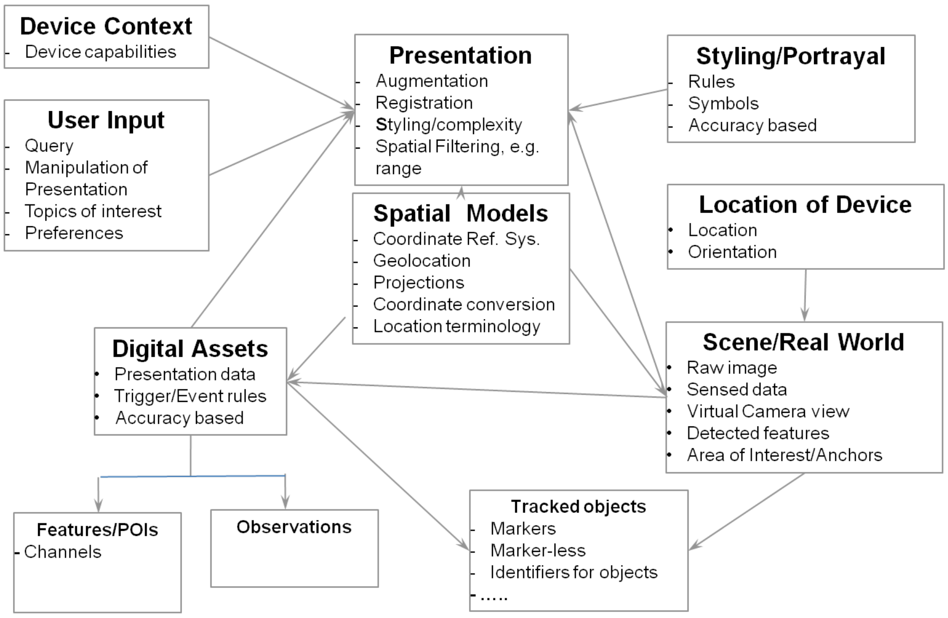
– Facilitates common understanding of AR content
- Topics in Viewpoint for AR
  - Information Content of Virtual Objects
  - Spatial Registration
  - Context Dependent Viewing

END NOTES

The Information viewpoint describes all the categories of data that (together) combine to result in the “presentation” of an augmentation registered with the target to the user.

In other words, presentation is the goal and can be experienced only as a direct consequence of other information processes.

Editor note: Flow of this section is confusing to me. Currently it begins with spatial model and ends with presentation. Should the information viewpoint not begin with presentation and work “outwards” to explain all the blocks?



3.4.1 Spatial Models

Spatial models are used in presentation to register augmentations. The spatial model consists of all those inputs that define the presentation (the “pose”) with respect to two positions: the user’s preferred sense for the augmentation (e.g., eyes, ears) and the target. Only by knowing these with 3 degrees of freedom (X, Y, Z), can the augmentation be correctly presented.

Scenes and Augmentations include spatial referencing information to be used in the registration process.

For augmentation to be realistic, there must be a complete spatial model; the system must capture the target pose and the user pose at every point in time. In other words, the system that is used for the spatial model is defined and consistent, but the real world (the scene) may change.

(Extend to spatial-temporal model to include "events")

The spatial model can use any coordinate reference system. For example, the CRS may be based on distance and angles from the camera. Alternatively, in some scenarios (AR TV??_, the CRS can be based on distance and angles between the target, the augmentation and the screen. The camera and screen are combined in the mobile device so, in the case of a smartphone as the point of view, the CRS is that of the device.

Frequently, in geospatial AR, the Global CRS is the default and, since the smartphones have GPS embedded and use the W3C Geolocation standard, this CRS is available outdoors.

For indoor AR use cases, the Global CRS may also be utilized, however, it will be a "derived" position, based on translation of another positioning method.

In mobile AR cases in which the user seeks augmentations about stationary geospatial points of interest, the orientation of the user is also required. The smartphone compass can provide the orientation to the application using standards, such as for example the W3C DeviceOrientation specification. Unfortunately, current compass technologies are highly variable and sensitive to environmental conditions and require frequent calibration.

Orientation can also be derived in other ways. For example, if one assumes that the user is oriented in the direction of travel, then this can be estimated solely on the basis of the geospatial coordinate at two points separated by a known time interval. The OGC published the MovingObjectSnapshot, an application schema of the OGC Geography Markup Language. The shema is proposed for use to encode a snapshot of a moving object including its location, translational velocity and acceleration. This specification addresses the use case of a solid object, such as a car, travelling in a plane local to the object, such as a street. The velocity is an instantaneous vector composed of a scalar speed and a heading relative to North.

3.4.2 Scene (Physical Object)

Video image of physical reality local to user.

GEORGE PERCIVALL 7/6/12 12:48 PM
Comment [17]: Add visual search, e.g., MPEG **N12201**

3.4.3 Augmentations

Editor's Note: there is a lot of material related to this topic in the glossary of terms. How to best integrate in the viewpoint?

Presentation Data (is this Digital Objects)
An augmentation is uses a digital object, information stored and transmitted in digital format and displayed to the user in real time and registered with the target or subject of the user's attention. There is no universal definition or recommendation for the size or format of a digital object used in Augmetned Reality.
Digital objects can be static, such as text, computer graphics, photographic images. Digital objects can also be dynamic such as animations, video or sounds. Digital objects should scale to match the size of the real world subject with which they or it is associated.

Trigger (see glossary)

- those attributes of the Point of Interest which (a) selected during authoring by the content creator and (b) when detected by the sensors of the user's system, serve to specify the presentation data to be used.

Editor's Note: this is where we say that a marker is just a type of trigger. There is no benefit to treating the topic of marker versus marker-less AR in any greater detail in this document.

Point of Interest (POIs)

- Individual data item associated with a geographic location (longitude, latitude, altitude) or a visual pattern (marker, book cover etc.) that can be rendered in some way by the AR application. (Burchart 2011)
- Augmented Reality browsers will need to integrate an extremely accurate POI database as well as geocoding functionality (NAVTEQ 2011)
- information about Point of Interest includes label (name), unique identifier and relative or absolute location. Additional information about a point of interest can include author, time created, time updated, description (category, or civic address). (W3C PoI WG)

Associated with a point of interest there also will be an area of interest which can be defined in any CRS and can be in any shape.

Channels, Layers and Worlds:

- All refer to published groupings of related POI's and associated virtual objects. Often channels provide access to all the content of a single content publisher (Butchart 2011)

3.4.4 Features

Editor's Note: Natural Features? Features, in any artificial intelligence context, are the result of analysis of observations? Or are we referring to the recognition of features performed as a result of comparing the observation (set of observations) that are acquired for the user?

3.4.5 Observations

Stimuli received/detected in real time by the sensors on the user's chosen device and evaluated in real time by the other AR system components in a fashion which permits detection of any trigger data in the user's environment.

The observations may be optical (from the camera), auditory (from the microphone), geospatial (as described above).

Observations may be produced by sensors on the user's device (e.g., smartphone), or by a remote sensor. For example, a video camera mounted in a fixed position and pointed at an area of interest to the user. This may be called the "indirect" AR use case.

Due to the real time nature of the experiences, observations must always associated with a time or time interval (timestamped) for Augmented Reality. If observations are received out of sequence, they may be deleted or used, depending on the system.

3.4.6 User Context

What Digital Data to Present to the user?

- Discovery (of user's environment based on observations)
- Filtering (OK but isn't that also discovery?)
- Matching trigger data with presentation data

Christine Perey 7/6/12 12:48 PM

Comment [18]: This is just at the edge of being clear and maybe could be interpreted as the

- Portrayal and symbols defined by the application in use
- Context

Christine Perey 7/6/12 12:48 PM
Comment [19]: Not sure, not sufficiently specific

Figure 2 – Context Driven Detail

Figure 3 – Search by Looking

3.4.7 Styling

- depending on the computational complexity and the power of the device, there may be several different options for the resolution and size of the digital data in the presentation layer.
(TBD)

3.5 *Computational Viewpoint for AR-RM*

Editor's note: Provide the scope of the computational viewpoint.

- Viewpoint definition

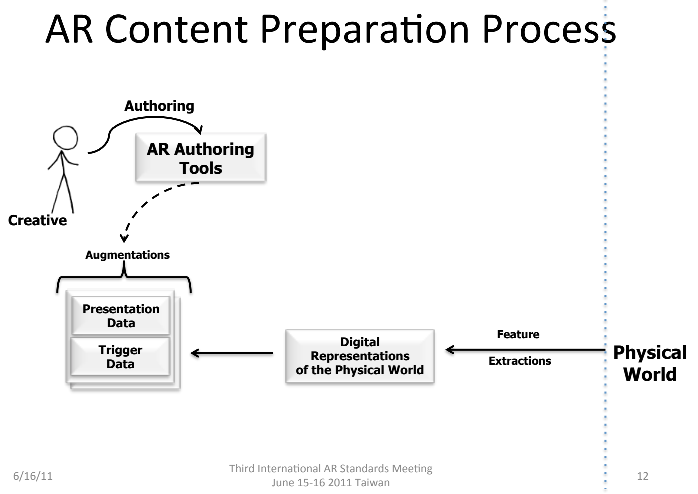
- Identify interfaces that allow for distribution
- Types of services and protocols
- *How does the bits work together and a part?*

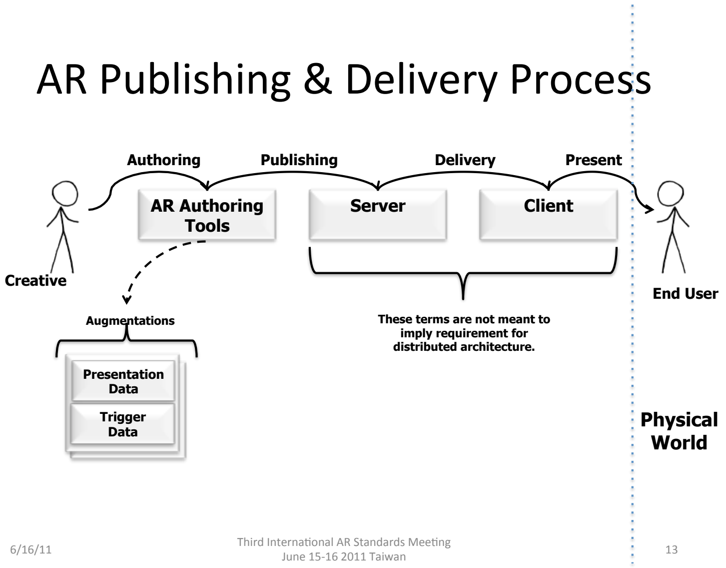
- Topics in Viewpoint for AR
  - *Use cases*
  - *Identify services and interfaces*

3.5.1 Use case: Content Authoring

Editor's note: There is no description on each block. Explain figures below.

Use case about content authoring automatically in bulk, by user and in other potentially interesting combinations.





3.5.2 Use case: Content Consumption

Editor's note: It would be better to have the overall service or functional description first.

3.5.2.1 Determine Location and orientation

- Determine location and orientation

Example Location determination

- Device, using GPS and compass or another spatial position detection technology (accelerometer, radio)
- Third party, using collected information
  - Observations: Wireless signals, ping/traceroute
  - Direct/purchased access to information about infrastructure
- Network operator, using network
  - Desired case for emergency services / 911
  - IETF GEOPRIV working group

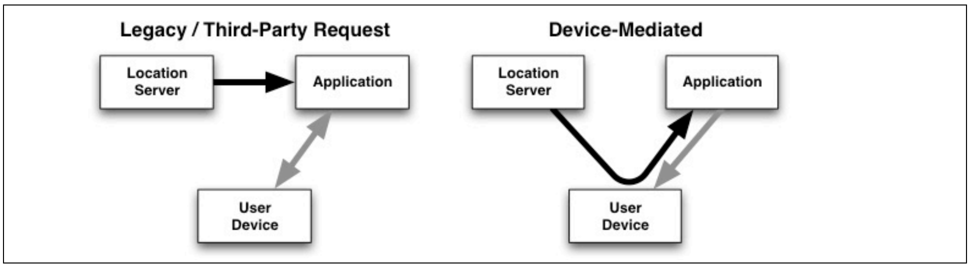


Figure 4 – Use Cases for Location Determination

Reference: Geolocation and Location-Based Applications [Barnes 2011]

–

3.5.2.2 Object Detection/Identification

- Processing of visual image from on-board camera to determine
- Sensor processing

3.5.2.3 Digital Asset Retrieval

- Retrieve local/remote DA based on geographic location
- Retrieve local/remote DA based on feature identification

3.5.2.4 Presentation

- Present visual digital assets registered to camera image
- (Extend to other senses: audio, haptic,)
- Create Content
- Content provider creates content and publishes

3.6 *Engineering Viewpoint for AR-RM*

- Viewpoint definition
  - Identification of component types
  - To support distributed interaction between the components.
  - How do the components work together?
- Topics in Viewpoint for AR
  - Engineering Components
  - Interfaces between components
  - Device Internal Interfaces
  - Encoding Formats

3.6.1 Internal Components and Interfaces between components

- Camera and Display
- Location and orientation determination
- Local physical environment sensing
- Persistent storage of content
- External communications including cellular, WiFi, Near-field
- AR content management software components

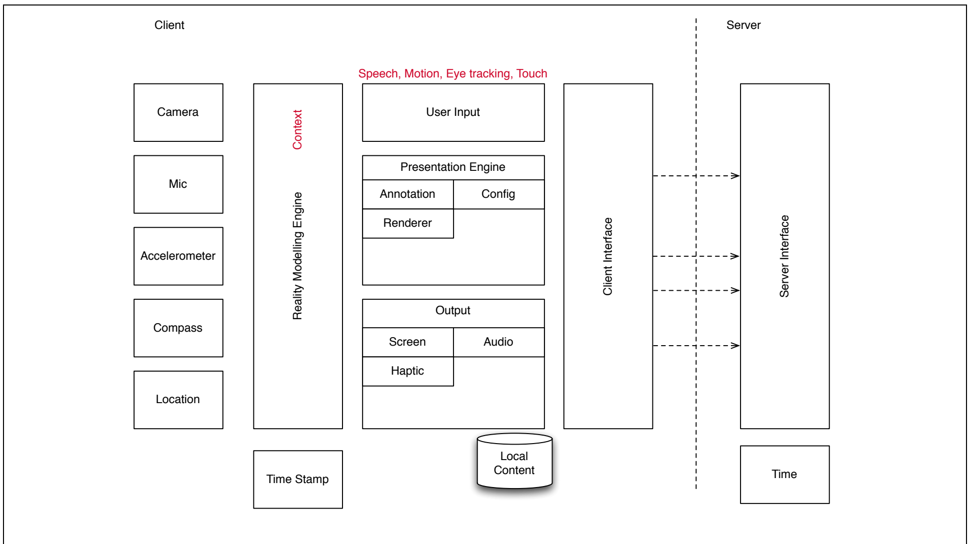
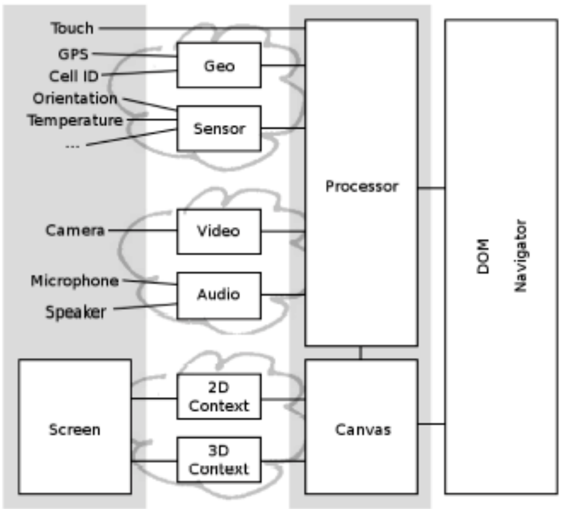
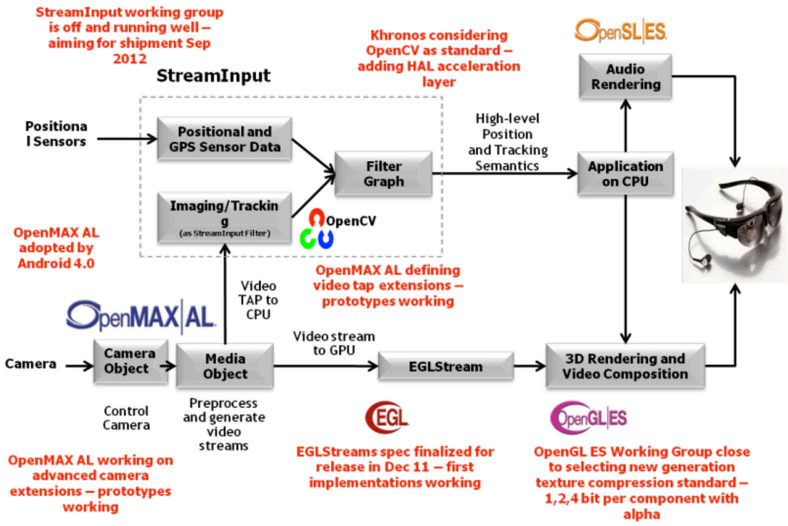
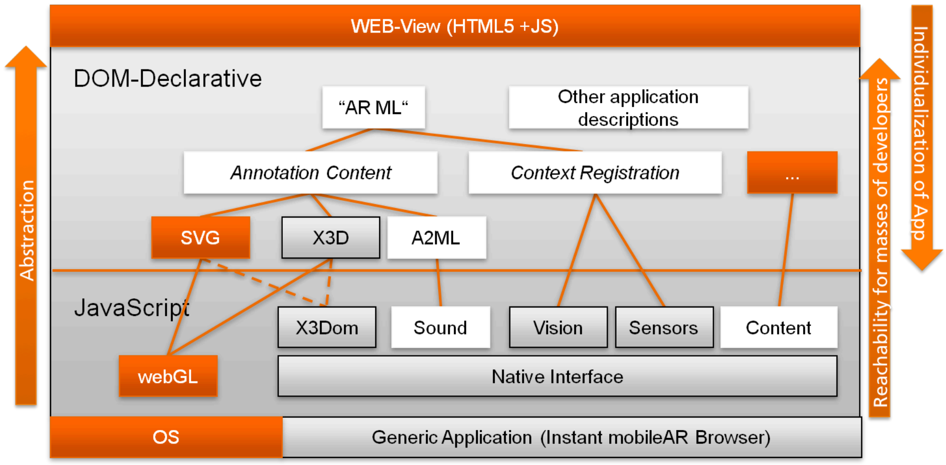


Figure 5 – Mobile Device Internal Components

Reference: 2nd International AR Standards Meeting, February 2011







3.7 *Technology Viewpoint for AR-RM*

Editor's note: There is no data available for this viewpoint.

4 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

4.1 *Animated Objects*

Animated Objects are digital objects in a composed scene that have been given (by the developer of the object) a dynamic property (changing over time). The behavior (or dynamic property) can begin upon appearance or be the result of user's interaction with the digital object. There are also digital objects (e.g., virtual humans) that can move within a scene in a fixed trajectory.

4.2 *Auditory AR Experience*

An **Auditory Augmented Reality experience** is the result of a system providing the user digital audio as a result of detecting some trigger in the user's proximity. The trigger may be visual (computer vision), auditory (natural language), geospatial or detection of other environmental conditions such as radio signals.

4.3 *Augmentation*

An **augmentation** is a relationship between the real world and a digital asset.

The realization of an **augmentation** is a composed scene.

An **augmentation** may be formalized through an authoring and publishing process where the relationship between real and virtual is defined and made discoverable.

Note: The term "augmentation" is sometimes used as a verb, used loosely to describe the composed scene and the related process of detecting the pose of the real world observer relative to a real world object (cf. registration) to realize a convincing rendering of digital asset in a real world setting.

4.4 *Augmented Reality Application*

An **AR application** or **AR app** is any service that provides augmentations to an AR-ready device or system. Sometimes it is useful to make a distinction between a single purpose (or narrowly themed) "app" and an "AR Browser" that can offer a user access to augmentations using the browser provider's content authoring system. In this distinction, an AR application would normally only consume content from a single source content provider or a restricted set of trusted providers whereas a "browser" would support content from many sources.

4.5 *Augmented Reality Browser*

The term **AR Browser** refers to a class of AR applications that offer a wide variety of AR experiences and themes from more than one content provider. Browser vendors will typically offer a publishing platform and will either host content themselves (in the browser provider's content management system) or offer a mechanism for others to host content that can be served to the browser on demand. At the moment, the distinction between an "AR Browser" and "AR App" is fairly loose, as the industry lacks standards required to implement compliant browser applications.

4.6 *Augmented Reality Content Management System/Platform*

An Augmented Reality **Content Management System** is a database with well defined interaction types that a content provider can use to produce AR Experiences. Augmented Reality CMS normally offer data hosting and provide a ***Graphical User Interface*** with the ability to specify locations for Points Of Interest on an map, upload and process reference images, add actions to digital assets and preview and publish the content. Some CMS offerings can prepare content to publish to several different AR Browsers. Sometimes the CMS is provided to support a particular AR Browser or application, allowing content providers to add digital assets and reality anchors to a database controlled by the application provider. Occasionally, the term CMS is used to refer to a ***Software Development Kit*** that enables developers to rapidly create a hosting environment on their own servers.

4.7 *Augmented Reality Experience*

An augmented reality experience (also an AR User Experience) is that which is produced as a direct result of combining, in real time, one or more elements of the physical world, one or more augmentations and related user interactions.

4.8 *Augmented Reality Marker*

An augmented reality marker is a 2D (frequently black and white and square in shape) symbol that looks like a 2D barcode and serves as a trigger for an augmentation. It is defined within an AR authoring platform and is unique for each augmentation. There is no defined standard for "AR Marker", however, many applications with remote or embedded computer vision algorithms are capable of recognizing an AR marker.

4.9 *Authoring*

Authoring is the process of creating a link (an augmentation) between a digital asset and the real world. The author must define how a digital asset will be rendered and how it is linked to a real world environment. The author can specify a reference object in the real world (Point of Interest , a reference image, or marker) to anchor the digital asset in a composed scene. Authoring may also involve the specification of behaviours that can apply to the digital asset. The concept of authoring differs from the term ***modelling*** which describes the creation of 3d scenes, although 3d modelling (or preview) might be incorporated into an authoring platform. The output of authoring is often some form of ***mark up*** which provides a structured format for describing the augmentation. Authoring also involves the specification of styling and formatting options. ***Categorization*** of virtual objects falls into authoring where the purpose is to provide a presentational ***filter***. The same categorization of digital assets can assist ***discovery*** and is therefore part of the publishing process. In the publishing context, categorization is sometimes called ***tagging***.

4.10 *Behaviour*

A **behaviour** is a feature of a digital asset that enables the user to manipulate or visualize the object in a variety of ways. Example behaviours include rotating the virtual objects relative to the users' point of view, animations, calling a phone number or opening a URL associated with the augmentation in a web browser. A behaviour can be activated by user interaction or by sensors. For example, a behaviour could be activated when a GPS sensor detects the user has arrived at a specified location or the camera (light sensor) detects a predefined pattern. Behaviours are often referred to as ***scripts .***

4.11 *Camera View*

Camera View is the term used to describe the presentation of information to the user (the augmentation) as an overlay on the camera display.

4.12 *Composed Scene*

A **composed scene** is produced by a system of sensors, displays and interfaces that creates a perception of reality where augmentations are integrated into the real world.

A **composed scene** in an augmented reality system is a manifestation of a real world environment and one or more rendered digital assets. It does not necessarily involve 3D objects or even visual rendering.

The acquisition of the user (or device) 's current pose is required to align the **composed scene** to the user's perspective.

Examples of **composed scenes** with visual rendering (AR in camera view) include a smartphone application that presents a visualisation through the handheld video display, or a webcam-based system where the real object and augmentation are displayed on a PC monitor.
Note: **A composed scen**e is different from the terminology used in 3D modelling, where the 3D scene simply describes camera angle and lighting but is often divorced entirely from a real world environment.

4.13 *Digital Asset / Digital Object / Virtual Object*

A **digital asset** is data that is used to augment users' perception of reality and encompasses various kinds of digital content such as text, image, 3d models, video, audio and haptic surfaces.
A **digital asset** is part of an augmentation and therefore is rendered in a composed scene.
A **digital asset** can be scripted with behaviours. These scripts can be integral to the object (for example, a GIF animation) or separate code artefacts (for example, browser ***mark up***).
A **digital asset** can have styling applied that changes its default appearance or presentation.
Digital assets are sometimes referred to as ***content,*** but this is more general in its use whereas digital assets are understood as components in an augmentation.
Note: A digital asset is normally understood as a single entity from the user perspective, even if it is technically composed of several artefacts. So ***textures, materials*** and ***scripts*** would be bundled together as part of the same object even if they are physically separate files. A digital asset is a broader concept than ***model*** as it incorporates a variety of content types - not just 3d models and scenes.

4.14 *Geo[spatial]-based Augmented Reality*

Geo- or **location** based AR refers to augmented reality experiences based on the user 's location and orientation in a geographic coordinate space. Therefore the registration and tracking system relies principally on geo positioning techniques. Most frequently, the user's position is approximated from the location of the user's device based on one or more sub systems such as GPS, WiFi or cellular geo positioning. Sometimes the user enters a location manually or scans a LLA marker. The user's orientation is approximated from the movement of the device using sensors such a digital compass, accelerometer and/or gyroscope. Together with a location fix, the orientation sensors can provide enough information to approximate the user's 6 degree of freedom pose. Geo-based registration sometimes provides a first approximation for obtaining a user pose which is then refined using computer vision techniques. Another class of system, called ***external tracking systems***, use external cameras to detect and track the position of a user relative to the known camera position.

4.15 *Haptic AR Experience*

A Haptic Augmented Reality experience is the result of a system providing the user a vibration, temperature change or introduction of another sign detectable by the user's sense of touch as a result of detecting some trigger in the user's proximity. The trigger may be visual (computer vision), auditory (natural language), geospatial or detection of other environmental conditions such as radio signals.

4.16 *Interaction*

Defines how users interact with digital assets, how augmentations are presented to the user, how the user can provide input to an augmentation, actions such as search and filtering that the user can perform. Behaviours are a subset of user interactions that relate to how the user interacts with digital assets. Interactions also describes how digital assets react to external events and changing condition in the real world. (i.e. event not initiated by users)

4.17 *List View*

List View is the term used to describe the presentation of relevant information to the user in a list organized by alphabetical order, relevance score, date or another filter.

4.18 *Longitude Latitude Altitude Marker*

A Longitude, Latitude Altitude Marker is a planar symbol, a particular type of AR Marker, that, when detected by an application which is in communication with a content management system, receives an absolute or relative user (device) position.

4.19 *Map View*

Map View is the term used to describe the presentation of relevant information to the user (the augmentation) in a geospatial coordinate system depicted on a map.

4.20 *Markup*

Markup is the use of encoding of augmentations, triggers and any other information to create the composed scene.

4.21 *Point of Interest*

A **Point Of Interest** (POI) provides a geospatial anchor for an augmentation. Typically the Point Of Interest specifies geographic coordinate (e.g. WGS84 longitude, latitude, altitude) or a set of points representing area of interest or other ***geographic feature***. The Point of Interest links the geographic location to the augmentation and is used in an extended sense to include metadata (address, feature type etc.), styling and behaviours. The term Point of Interest is in common use by geographers to link a spatial geometry to any kind of geo referenced data, not just to augmentations. There is a close similarity in function between a Point Of Interest and a reality object in that they both provide the "reality" part of an augmentation. Sometimes the terms ***Feature of Interest*** or ***anchor*** are used to capture both geographic and reality objects in a single term. It is not uncommon for developer to use the acronym ***POI*** to describe any anchoring of data to a real world, even if the link is not a geographic feature. In this use ***POI*** [pronounced "po-y"] is used as a synonym for an augmentation.

4.22 *Pose / Six Degrees of Freedom Pose*

A real object in space can have three components of translation - up and down (z), left and right (x) and forward and backward (y) and three components of rotation - Pitch, Roll and Yaw. Hence the real object has six degrees of freedom.

4.23 *Provenance*

The **provenance** of a virtual object used in an augmentation is its source prior to the use in an authoring / publishing process or use in an end user AR experience. An augmentation's provenance can include information about the content creator, the date of the virtual object's publication or other metadata.

4.24 *Publishing*

Publishing enables an augmentation to be discovered. This includes the provision of metadata, the formatting of digital assets and the transfer of data to one or more servers to make the link discoverable by search engines, crawlers and AR clients. Publishing is closely linked to authoring and often the two processes are supported by the same content management system.

Authoring focuses on the creation of augmentations whereas publishing concerns the ***discoverability*** of augmentations. The term ***search*** describes the user interfaces and APIs that are used to discover both the augmentation and related metadata. The term ***filtering*** applies to the presentation and styling of information so is usually related to authoring rather than publishing.

4.25 *Real Object / marker / anchor*

A **real object** or a **marker** is a feature or artefact in the real world that is used to **anchor** an augmentation in a composed scene. This includes ***natural features*** such as buildings and landmarks, artefacts such as posters, book covers or pictures, and **markers** such as barcodes, 2d matrix codes and other machine readable patterns. A distinction (***marker vs. marker-less***) is frequently made between objects that are part of the everyday environment (posters, book covers etc.) and objects that have been created specifically for the purpose of an augmentation (markers, barcodes). In computer vision, the term ***feature*** is used to describe a pattern that image recognition algorithms can use to identify an object. This ***feature extraction*** class of algorithms is central to image recognition approaches to registration and tracking. Typically a **real object** would have several distinctive features that detection algorithms can identify. A **real object** is sometimes called a ***target, anchor*** or trigger [computer vision]. A real object is usually represented internally in the system as a reference image – a digital image with additional metadata or formatting that assists feature extraction.

4.26 *Reference image / reference object*

A reference image is a representation of a real object (usually an image of the object) that is used by image recognition algorithms to match a frame from the composed scene so that an anchor point for the augmentation can be identified. Often the authoring process creates a version of the original image encoded in a format more efficient for the image processing algorithms. In the context of a visual search, a reference images refers to one of the images used by the search engine as a ***search index***.

4.27 *Registered Scene*

A registered scene is a representation captured after the registration system has detected a real object and is used subsequently as a reference point for tracking. This registered representation is optimized for tracking algorithms and might differ from composed (rendered) scene presented to the user.

4.28 *Registration / Detection*

Originating from ***computer vision***, the term **registration** (also known as ***Detection***) – describes a system for providing an initial 6 degree of freedom ***pose*** relative to a real object or previously registered environment. The pose usually represents the physical location and orientation of the viewing device or its camera relative to a known point or object. The registration system can obtain a pose using either sensor data or/and computer vision techniques. In the case of computer vision, the registration system will involve a classification step, where a visual search operation detects a **reality object** using a set of pre-defined **reference objects**. In sensor based registration, sometimes called ***location based registration***, the geo positioning sensor is used along with orientation sensors such as a magnetic compass, gyroscope and accelerometer to approximate an initial 6 degree of freedom pose. It is common to combine both vision and location based registration techniques in a single registration system.

4.29 *Styling*

Styling specifies how the digital asset will render in the composed scene. This includes the specification of colours and fonts, specifying the size of symbols, defining what symbols should be used to represent different categories. Styling is an optional part of the process of authoring that overrides the default rendering of an augmentation. Often styling options are incorporated into the same ***mark up*** that describes the augmentation. However, a style sheet can be logically and cleanly separated from structural content.

4.30 *Tracking*

Tracking describes a subsystem for providing a 6 degree of freedom pose relative to previously ***registered*** real object or previously registered ***scene***. In contrast with registration, **tracking** uses the previously known pose for generating a new one based upon the frame or location fix that came before. Typically, computer vision algorithms exploit ***features*** extracted from the reference object to perform tracking. The system can either use previously extracted features or generate ***features*** from the reference object on-the-fly. Where a system's registration process is too slow to be used in frame-to-frame mode, tracking is unavailable or sporadic. If an algorithm is fast enough to register an object in real time at a reasonable frequency (i.e. 25Hz (*reference?*)) the method usually is called ***tracking by detection***. In the case of location based tracking, the tracking system often obtains a new location fix based on the last, using delta measurements taken from fast, low power sensors.

4.31 *Trigger*

A **trigger** is the condition(s) that will cause an augmentation to be sent to the user. When a reality object is detected, and the trigger conditions are met, a system pushes one or more augmentations and any associated interactions to the composed scene. In computer vision, the term **trigger** often refers to the salient attributes of the real world object or marker (or sound) that are necessary to facilitate rapid detection. As a result of a match with a **trigger**, the digital asset, and any embedded or associated interactions, is rendered by the device output and display system (including visual, haptic or auditory experience). The term trigger is a common coinage in computer vision and is close in use and meaning to terms such as reality object or real world object, but specifically emphasizes the representations used for computer vision techniques and the actions and behaviours that a visual match will generate.

4.32 *User Pose*

The **User Pose** is the position of the user relative to the target for augmentation and other parts of the real or digital environment. It can be used as an input to registration and also as input to a composed scene.

4.33 *User Query*

A **User Query** is a user-driven (user initiated) request for a digital object or digital asset as part of the AR Experience. It may be communicated by the user via gaze, touch, pointing, speech, text or another input method.

4.34

Visual AR Experience

A Visual Augmented Reality experience is the result of a system providing the user a digital object displayed in the user's field of view in response to detection of one or more triggers in the

user's proximity. The trigger may be visual (computer vision), auditory (natural language), geospatial or detection of other environmental conditions such as radio signals.

4.35

Visual Search

Visual Search involves obtaining information about a real world artefact by submitting a digital image or any subset of the image to a ***visual search engine***. Visual search does not, by itself, constitute AR. It can be very valuable detecting the trigger for an augmentation. Clearly a link is being made between the real world and some digital content but the user experience does not greatly enhance of users' perceptions of reality (in real time) so it is not clear if a composed scene is part of a visual search experience. Visual search can be used for discovering augmentations, for example, a barcode, 2d matrix code, logo or package design could be used to discover an augmentation applied to a box of cornflakes. In ***computer vision***, Visual Search is typically used for classification of an object or retrieval of information to be used for further registration and tracking.

5 Use cases

Editor's note: The use cases are obtained from the output document from San Jose meeting and the two input documents from Geneve's meeting. In addition, the entire Subclause is copied from "the Mobile Augmented Reality Use Cases for Standards Development Uses". It needs to be edited to fit into this document.

5.1 *Introduction*

The AR (Augmented Reality) is a new trend in various service areas such as advertisement, entertainment, education, and etc. in platforms such as mobile, tablet, PC, and TV. The 2D/3D graphics is integrated into the real world in order to enhance user experience and enrich information. Figure 1 shows a classic example of virtual advertisement in sports broadcasts. They usually use the sports ground or crowd as a place to overlay ads. This case is to do all augmentation before the video image is transmitted. Once it is transmitted, the viewer at home may never influence this compositing, no interaction for the user is possible unless the viewer has a direct connection to the content producer.

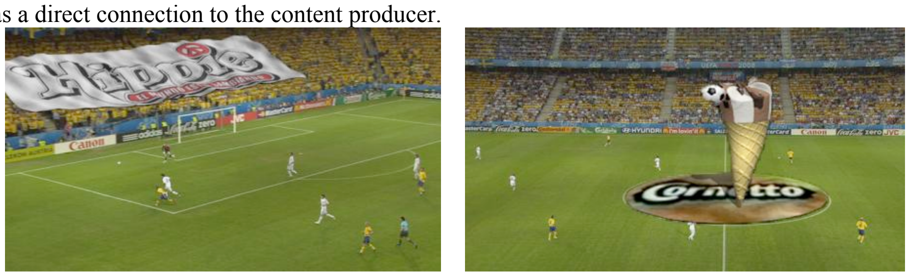


Figure 6 – Virtual advertisement in sports broadcasting

Christine Perey 7/22/12 11:53 AM
Deleted: 1

AR becomes more attractive when interactivity is allowed. Mobile device provides user interface such as touch screen, and various sensors such as motion sensors, and GPS sensor by which more interactive AR service can be implemented. Mobile as a hand held device also allows user to walk around in the displayed scene at free will and composites useful information related with the location of the user. On the other hand, TV is usually one way service and has restriction on

the interactivity. The free movement and thus the free selection of the angle towards the live scene are unfortunately not possible in a TV environment. However, as the Smart TV is getting popular, the internet connection is mandatory and a remote controller equipped with motion sensor enhances the user interface. Figure 2 shows an AR example with interactivity. Users can manipulate 3D objects according to the movement of marker image. Although this interactive AR is not real in the broadcasting environment yet, we expect to see this kind of service in TV soon.

Figure 7 – AR service in education

Christine Perey 7/22/12 11:53 AM
Deleted: 2

In this contribution, we propose use cases for AR blended within TV environment and requirements to support them.

5.2 *Augmentation region based AR*

The augmentation region is a coordinate space in the video image that 3D objects shall be overlaid. The 3D objects shall be displayed synchronized with the movement of the augmenting region. Figure 3 shows an example of augmenting region based AR. The augmenting region and its tracking information should be described beforehand because current set-top-box may not have enough power to process tracking automatically from the video stream in real time. Therefore it is necessary to prepare the trajectory of the movement and transmit the trajectory together with the video data. More over it would be possible to transmit a prepared mask, depth-information, 3D object, or other additional data such as illumination resource for natural composition of 3D objects.

Figure 8 – Augmentation region based AR broadcasting

Christine Perey 7/22/12 11:53 AM
Deleted: 3

5.3 *Interactive AR with manipulation of the 3D object*

Enabling a user to manipulate 3D object augmented over the background video will enrich the user experience. For example a student is watching TV program about the organ of the human body. During a teacher is explaining about the functionality of each part of the human heart, 3D model of heart will show up and the student may manipulate the 3D object by handling remote controller prepared with motion sensor. Another option to manipulate 3D object is to run image

recognition algorithms for tracking of the marker image. This will be possible if the cost of the set-top-box drops and more calculation power and CPU speed are available to the set top box in the near future.

Figure 9 – AR with manipulation of the 3D object

Christine Perey 7/22/12 11:53 AM
Deleted: 4

5.4 *Choice of AR service provider*

As the users select their prefer TV channel among various broadcasting service providers, the source of AR also could be chosen by the user. This is not realized in TV environment yet however similar concept is realized in mobile environment already. Example service is “The Layar Reality Browser“ which provides various types of AR services such as location-based layers help users to find nearby locations, including cafes, shops and other businesses, as well as historical locations and monuments. Other layers let users play games within their environment, browse for clothes in a 360-degree virtual shop, or even artwork placed digitally into the real world. Users can select one of those layer services working on the Layar platform. Figure 5 shows Layar application which allow user to change layers (left image) and displays searched result based on the selected layer (right image).

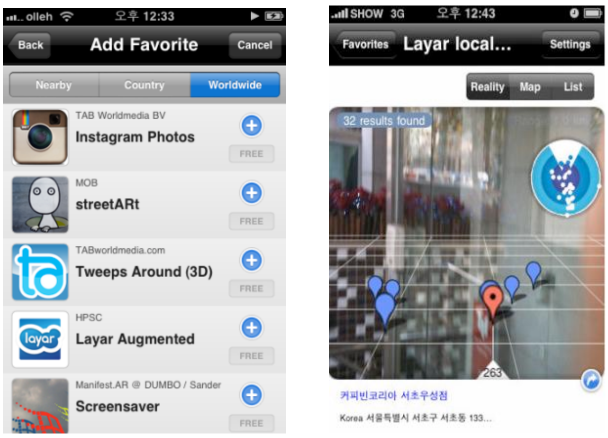


Figure 10 – Layar application

Christine Perey 7/22/12 11:53 AM
Deleted: 5

In the TV environment, it is also possible to provide this kind of service. The broadcasters transmit video with the augmentation region, time, and location information but not include the actual overlay objects. Instead they provide the list of access points of AR service provider by which the users can select the AR service provider and access their overlay objects. This business model will reduce the overload to develop AR platform for each application, bring opportunity to the potential AR service providers, and provide the users the options to select qualified AR service among the various service providers.

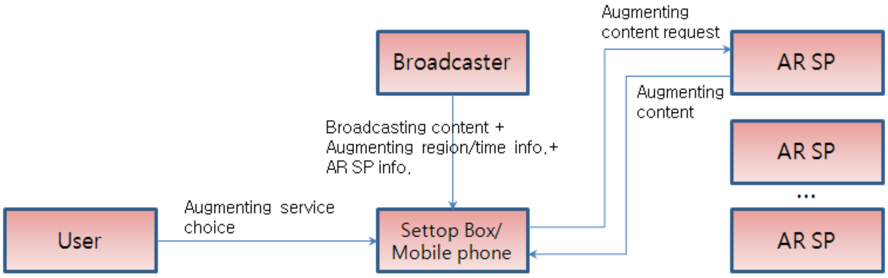


Figure 11 – Choice of AR service provider

Christine Perey 7/22/12 11:53 AM
Deleted: 6

5.5 *Location based AR broadcasts*

Location based AR services are very popular in mobile device nowadays. Mobile devices usually have location sensor and orientation sensor. However TV does not have such sensor and the video stream is not related with the location of user's watching point. Then what if the video stream contains the location and orientation signal obtained from the camera? TV camera equipped with location sensor and orientation sensor is already in used in AR advertisement in sports game broadcasts. Not only the advertisement, but various useful information based on the location could be possible. For example, in the golf game broadcasts with location information, the AR service provider may give the users the ground shape of the field. Another service provider gives the information about the golf club. Similarly in a tour broadcasts, location information allow user to more useful information such as transportation guidance or some restaurants nearby the place in the scene.

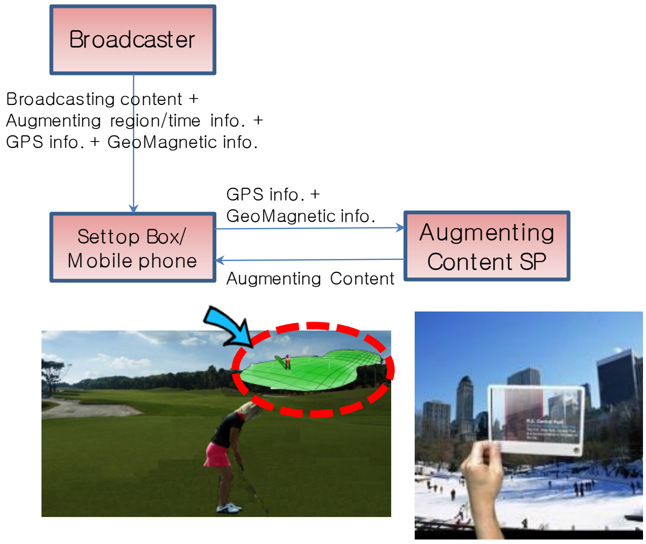


Figure 12 – Location based AR broadcasts

Christine Perey 7/22/12 11:53 AM
Deleted: 7

5.6 *Medical Augmented Reality*

Doctors or physicians could use Augmented Reality (AR) as a visualization and training aid for medical diagnosis or health education. It may be possible to collect the 3D anatomical data and health-related information of a patient, using non-invasive sensors such as Magnetic Resonance Imaging (MRI), Computed Tomography scans (CT), ultrasound imaging, sphygmomanometer, or thermometer. These data could then be rendered and combined in real time with a view of the real patient. In effect, this would give a doctor "an anatomical vision" inside the patient (see Figure 8).

Figure 13: Medical Augmented Reality

Christine Perey 7/22/12 11:53 AM
Deleted: 8

Applications for medical diagnosis or surgery should communicate with bio-sensors, which detect vital signals such as heartbeat, blood pressure, temperature. Thus, these applications may support intuitively medical health monitoring for patients. Such information also can be used for 3D virtual organ motion rendering.

Furthermore, medical AR applications may receive the 3D dense objects, which represent patient's organs. For realistic modeling and rendering, virtual organ data may consist of several millions of triangles, high resolution textures and animation. However, these data cannot be transmitted in real time under the limited network resources. Thus, the compression method for

3D object should provide efficient RD (Rate vs. Distortion) performance and spatial/quality scalability.

In addition, these applications may support the 3D compact description of a real world image/video for the interaction (including the photo-realistic rendering) between virtual objects and the real world. The 3D scene description should have compactness, simplicity, and stability of representation for the real-time application.

The overall process for the medical augmented reality can be summarized as follows (see Figure 9):

1. Sensors detect the medical information and send it to an AR system
2. Devices such as camera capture the real world image/video
3. A server for compressed 3D graphics data send virtual organ data including animation
4. The AR system creates 3D compact scene description from the real world image/video
5. The AR system decompresses the compressed virtual organ data.
6. With the 3D compact scene description, the AR system renders the decompressed virtual organ realistically.
7. Using the AR system, users can interact with virtual organ and create events.

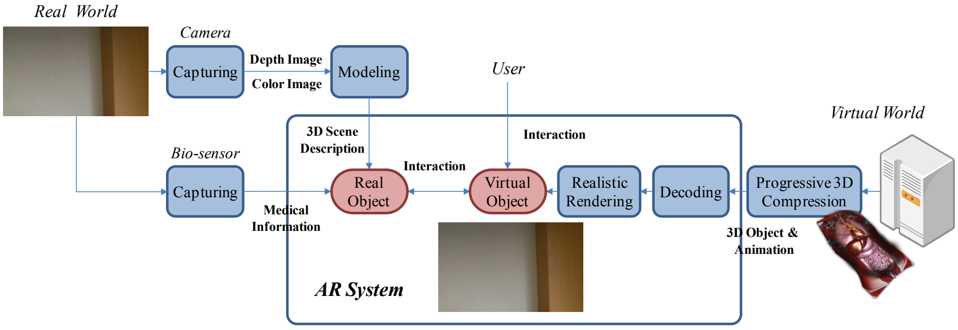


Figure 14: Overall process of medical augmented reality system

Christine Perey 7/22/12 11:53 AM
Deleted: 9

5.7 *Mobile Mixed Reality/Augmented Reality Games use case*

In the gaming area, AR/MR offers unique opportunities. Unlike with traditional video games, these games are not imprisoned onto the screen but interact with the real world. They incorporate real locations and objects into the game, therefore tapping into already existing thoughts and emotions of the player, which in turn creates a potentially much richer gaming world and experience. Mobile AR/MR Gaming even goes a step further as the playing area becomes borderless and games can be played literally anywhere and anytime. With the current

advancement of modern cell phone technology, we are identifying the potential of Mobile AR/MR Games being played by anyone.

There is no unique use case of MR/AR games, each kind of game being a use case by itself. We are providing here an analysis of the recent MR/AR games indicating the associated reference. Figure 10 illustrates some examples of games in MR/AR.

Figure 15: MR/AR Game examples.

Christine Perey 7/22/12 11:53 AM
Deleted: 9

5.7.1 Tidy City

5.7.1.1 Overview

Tidy City is a slow-paced location-based game for single players. In the game, riddles need to be collected at real world locations and brought to the real world locations they allude to.

5.7.1.2 Example of Play

Clarissa is walking around the Old Town of Cologne. She checks her smartphone running Tidy City, shakes her head and then looks around. She has just picked up the last riddle in front of the cathedral and cannot make anything out of it. “Love is a bridge, and Niklas and Verena are no exception, but theirs was just beginning.” In addition to this clue, her smartphone also displays an image of a lock to her with both names engraved (Figure 1, left). If Clarissa manages to solve this riddle and find the right spot, she will have solved the Old Town Cologne mission. If she just had an idea…

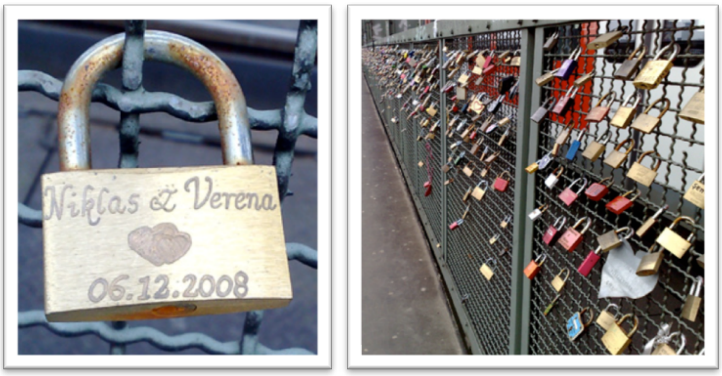


Figure 16: An image used as a clue (left) and the real world location for solving the riddle (right).

Christine Perey 7/22/12 11:53 AM
Deleted: 1

Clarissa continues walking around the area still thinking about the clue. She arrives at the banks of the river Rhine and notices a bridge in the distance. Maybe the “love is a bridge” part of the clue refers to a real bridge! Quickly she walks towards the bridge and when she is starting to cross it she suddenly sees hundreds of similar locking locks attached to the railing (Figure 1, right). Somewhere there must be the lock she is looking for! However, finding the right lock amidst all these others will probably not be too easy, so she takes another look at the clue: “but theirs was just beginning”. Maybe this part does not refer to their love but instead to the bridge? This would mean the lock is somewhere near the beginning of the bridge… And indeed, right at the start she finds Niklas and Verena. Clarissa presses the “solve” button for the riddle and is rewarded with points and a notification that she has finished the mission. She checks the highscore list and satisfyingly sees herself on top. Maybe she will now create a mission herself?

5.7.1.3 Detailed Description

When starting the game and after logging in, the player selects a suitable mission. The missions are displayed on a map showing their location. A mission has a name and a short description. After the player has chosen a mission all necessary data (including images) is downloaded to the device.

When the game starts, the player is presented with a map of the surroundings. It is possible to zoom in and out, move the map around, switch the layout between satellite imagery and stylized graphical representation and the current position of the player is displayed (Figure 2, left). Depending on the mission chosen, orange icons are displayed on the map. These icons represent the riddles. When the player taps on such an icon, more information about the riddle is displayed. When the player is still far away this information only consists of the name, the difficulty level (very easy, easy, normal, hard, very hard) and the category of the riddle (building, monument, nature, place, shop, restaurant, leisure, virtual object, other). The category is also reflected in the icon itself, so that there are nine different types of icons.

If a player is inside a 10m radius to the riddle (measured via GPS), the riddles displays additional information: a descriptive text and an image (Figure 2, middle). The player can also now pick-up the riddle and add it to his inventory which will gain him 1 point.

By combining clues from the different information provided by the riddle, it is now the task of the player to find the real world location where the riddle belongs. The player is free to collect additional riddles at all times during the game, i.e. there is no forced order of play.

Figure 17: Map view showing riddles and the player's location (left); information displayed when viewing a riddle (middle); player at the correct position for solving the riddle (right).

Christine Perey 7/22/12 11:53 AM
Deleted: 2

When the player is at a location that he considers being the right one, he can try to solve the riddle by pressing a button (Figure 2, right). GPS is again used to evaluate whether the player is close enough to the correct location (10m radius). If correct, the player gains points for solving the riddle depending on its difficulty level (5, 10, 20, 50, 100, 200). If not correct, the player loses 1 point, but can try again to solve the riddle at any time.

After collecting a riddle it disappears from the map view and is added to the player's inventory which he can access at any time. When a riddle is successfully solved, a light green icon appears on the map marking the correct position of the riddle.

The player also has access to a high score list which displays all scores by all players that have played (or are currently playing) the current mission. Anytime a player interacts with a riddle in a way that affects the score, the new score is uploaded to a server, so that it can be relayed to other players of the game.

5.7.1.4 The detailed use case

5.7.1.4.1 *Create a Tidy City Mission*

Actors

Actor	**Description**
MyMultimediaWorld Authoring tool interface	The authoring tool must be available via the website mymultimediaworld
Browser	Interface used by the game mission creator to create the mission
Game mission creator	Person (can be the player or not) that will create the mission, that is the play instance.

Initiator

Game mission creator

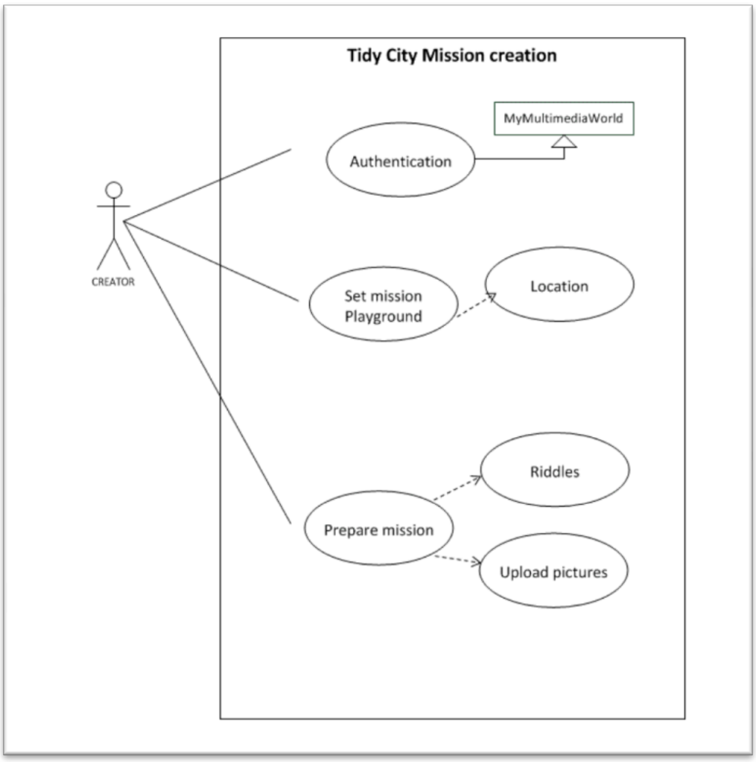
Preconditions

- Tidy City Authoring Tool is available
- Authentications is made in MyMultimediaWorld.com

Main Flow

Use Case 3		
Name: **User creates a Tidy City mission**		
Step	**System Actor Action**	**System Response**
1	Creator inspects the surroundings	No interaction with the
2	Creator takes the needed resources that are tideup with a location and creates the riddles	Pictures are saved in mymutlimediaworld
3	Creator access mymultimediaworld game creation page	The system asks for identification
2	Creator authenticates itself	The system goes to Authoring tool main page
3	Creator access Tidy City authoring tool	System displays current missions and offers possibility of new mission
4a	Creators creates a new mission	System allows the edition
4b	Creator access existing mission	System display the mission page with its characteristics

5	Creator edits mission	System displays the editing mode
6	Creator defines the object associating it to a category, defines the difficulty, the riddle, the “wrong” location and the solution (location). Note: There is no limitation to the number of objects	The location of the riddle and its solution is place on the map importing the asset information. The system saves the mission



TC game creation use case diagram

5.7.1.4.2 *Playing the Tidy City Game*

Actors

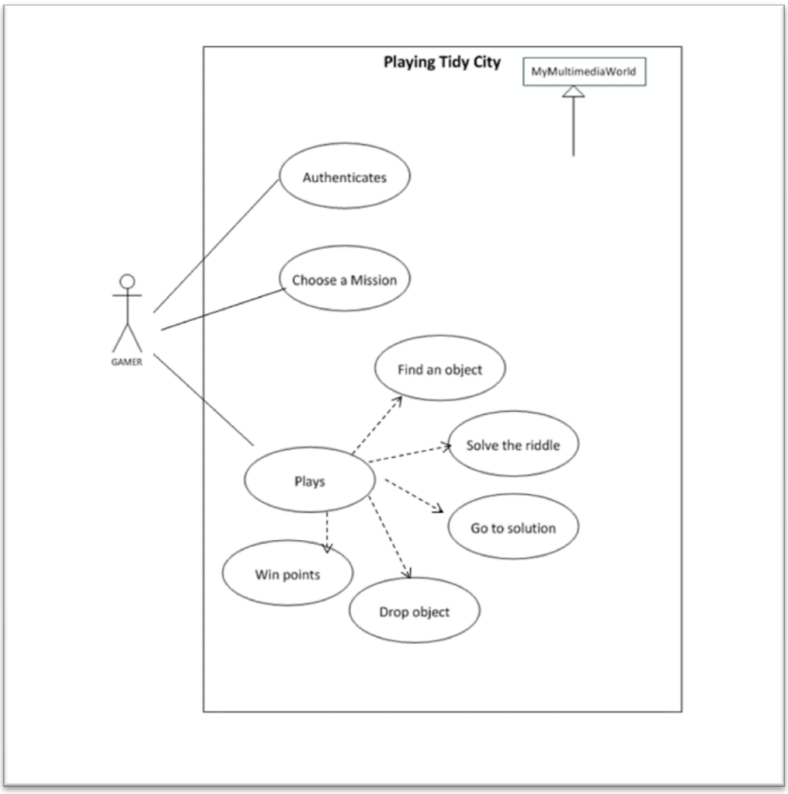
Actor	**Description**
MyMultimediaWorld Authoring tool interface	The authoring tool must be available and working for the
Browser	Interface used by the game mission creator to create the mission
Gamer	Person playing the game

Initiator
Gamer

Preconditions
- Authoring Tool is available
- Authentications is made in MyMultimediaWorld.com
- Gamer has the ARAF player installed in his Smartphone
- The Smartphone has GPS integrated

Main Flow

Use Case 4		
Name: **User plays a Tidy City mission**		
Step	**System Actor Action**	**System Response**
1	Gamer access mobile phone ARAF player	System ask for authentication
2	Gamer authenticates himself	The system connects to the server and confirm authentication Success: The system give access to AR games Failure: The system ask to restart
3	Gamer selects a mission	The system go to the mission page
4	Gamer inspects the surroundings and encounter a misplaced object	The system display the riddle
5	Based on the answer to the riddle, the Gamer solve the riddle and goes to the good location.	The System recognizes the location and a pop up window appear asking if the gamer wants to drop the object. If the position is the one corresponding to the object it add a point



Playing TC game use case diagram

5.7.1.5 Acknowledgement

The Tidy City game design was conceived by Michael Straeubig.

5.7.1.6 References

Wetzel, R., Blum, L., Feng, F., Oppermann, L. and Straeubig, M.: Tidy City: A location-based game for city exploration based on user-created content. Mensch & Computer 2011

Wetzel, R., Blum, L. and Oppermann, L.: Tidy City – A location-based game supported by in-situ and web-based authoring tools to enable user-created content. Foundations in Digital Games 2012 (to appear)

5.7.2 Portal Hunt

5.7.2.1 Overview

Portal Hunt is a fast-paced location-based augmented reality game for teams of players. The players encounter so-called portals in the environment that they need to collect (see Figure 3). The portals have different movement patterns which makes this process more difficult. In addition, some portal can only be collected when all members of the team co-operate.

5.7.2.2 Example of Play

Team Red consisting of Miriam, Paul and Daniela has tracked down some portals in a small park and are checking them out in the augmented reality view. At first the three portals they see seem to be stationary but then they suddenly disappear only to reappear a few seconds later at a different location. By observing them a while longer they notice that the portals disappear every 15 seconds which does not give them a lot of time to capture them. The first two portals are caught rather easily however as they only require one player to physically walk into the portal to collect it. The last one is trickier as all three of them need to surround the portal and then at more or less the same time need to activate the capture mode on their devices. The first attempt fails as Paul did not manage to get fast enough to the other side of the portal, but the second try is successful.

Figure 18. Conceptual image showing one team approaching a portal.

Christine Perey 7/22/12 11:53 AM
Deleted: 3

Team Blue in the meantime has spotted another portal but it is seemingly floating above a lake. Jane, Barbara and Kim are unsure on how to proceed. They certainly will not swim towards the portal as that seems too dangerous for their smartphones. Kim scans the environment and finds a small rowing boat hidden behind some reed. It seems like all of them have to get on board as this type of portal can only be caught when all players enter it at the same time. To make matters worse, the portal is also not standing still but slowly moving around the lake. The three of them start rowing and by analyzing the movement pattern of the portal they manage to get the boat to a spot which the portal regularly passes. And indeed, they manage to catch it!

5.7.2.3 Detailed Description

In order to play the game, players need to join a game session and a team. When the game has started, the players can see a map view on their devices which shows the position of any portals in the vicinity. In addition to that the players can also switch to a camera view, to show an augmented reality representation of the portals. Positioning is done via GPS.
Portals have different characteristics that apply to their movement behavior and the method that has to be used to successfully catch them. They also have a point value assigned that is added (or subtracted) to the team's score when the portal is caught.

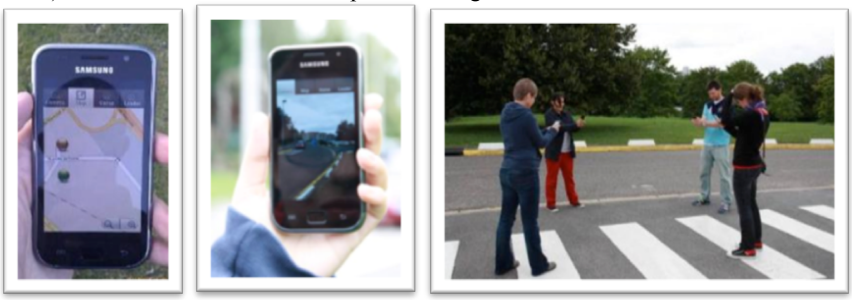


Figure 19: Map view showing the location of portals (left); portal displayed when viewing in AR (middle); players surrounding a portal (right).

Christine Perey 7/22/12 11:53 AM
Deleted: 4

The movement characteristics are:

- Stationary: The portal always stays at a fixed position
- Moving: The portal moves along a predefined path at a predefined speed. The path is a closed loop.
- Teleporting: The portal stays at a fixed position for a specific amount of time and instantaneously moves to another position (and then another etc.). After the last specified position the portal teleports back to the first position and the cycle repeats itself

The catching characteristics are:

- OneInside: Any one player of a team needs to walk inside a portal to catch it. These portals are colored green.
- AllInside: All players of a team need to be inside a portal at the same time to catch it. These portals are colored yellow.
- Surround: All players of a team need to position themselves around the portal in equal distance to each other and maximum distance from the portal. They all need to trigger the catch by pressing a button almost simultaneously. These portals are colored red.
- NoCatch: These portals cannot be caught. Instead, if any player comes too close, the team loses a certain amount of points. These portals are colored black.

The players also have access to an overview screen that shows the current points of all teams participating in the session. If all portals have been caught (or a time limit is reached) the game ends.

5.7.2.4 The detailed use cases

5.7.2.4.1 *Create a Portal Hunt Mission*

Actors

Actor	Description
MyMultimediaWorld Authoring tool interface	The authoring tool must be available via the website mymultimediaworld
Browser	Interface used by the game mission creator to create the mission
Game mission creator	Person (can be the gamer or not) that will create the mission, that is the play instance.

Initiator

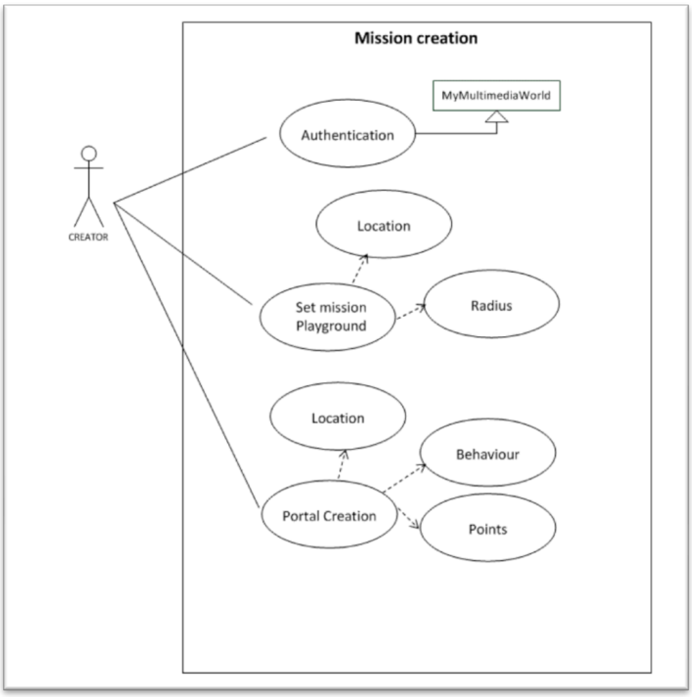
Game mission creator

Preconditions

- Portal Hunt Authoring Tool is available
- Authentications is made in MyMultimediaWorld.com

Main Flow

Use Case 1		
Name: **User creates a Portal Hunt mission**		
Step	**System Actor Action**	**System Response**
1	Creator access mymultimediaworld game creation page	System asks for identification
2	Creator authenticates itself	System goes to Authoring tool main page
3	Creator access Portal Hunt authoring tool	System displays current missions and offers possibility of new mission
4a	Creators creates a new mission giving the name and the description	System allows the edition
4b	Creator access existing mission	System display the mission page with its characteristics
5	Creator edits mission	System allows the edition
6	Creator choose the game zone (playground) by clicking on the map what corresponds to the center of the zone, the radius going to 1km to 10km, and the name of the mission. Note: The Creator can "navigate" in the map to find the place corresponding to the game. The Creator gives a name to the object and determine its behavior	The System displays the map is display and allows saves the zone
7	The Creator access portal creates the different elements (that is the portals).	System allows edition and save the changes



PH game creation use case diagram

5.7.2.4.2 *Playing the Portal Hunt Game*

Actors

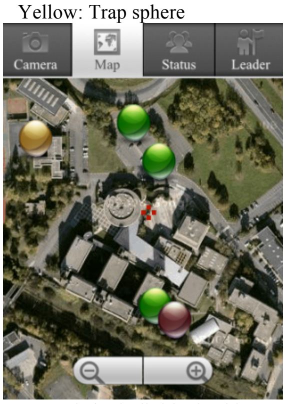
Actor	Description
MyMultimediaWorld Authoring tool interface	The authoring tool must be available and working for the
Browser	Interface used by the game mission creator to create the mission
Game mission creator	Creator of the mission (can be the gamer)
Gamer	Person playing the game

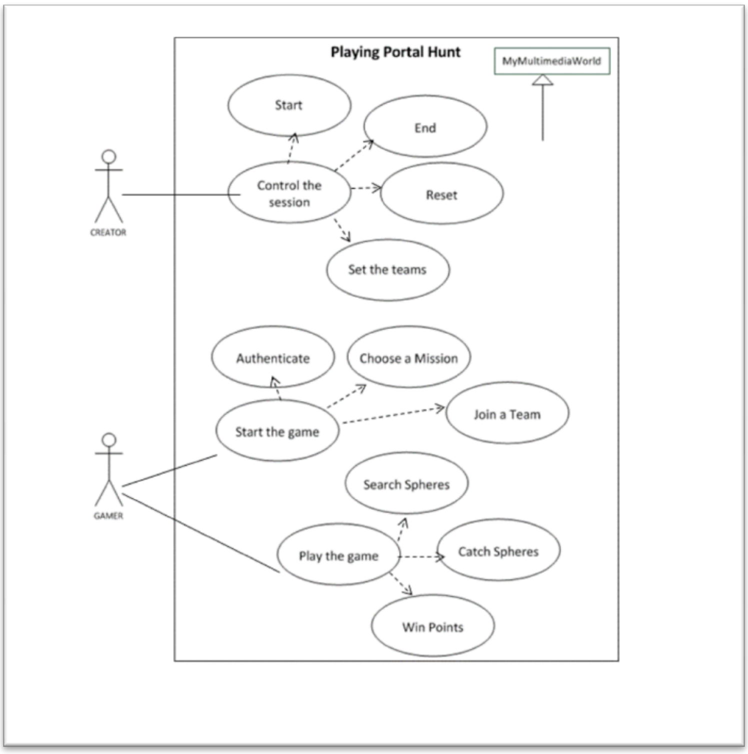
Initiator
Game mission creator / Gamer

Preconditions
- Authoring Tool is available
- Authentications is made in MyMultimediaWorld.com
- Gamer has the ARAF player installed in his Smartphone

- The Smartphone has GPS integrated

Main Flow

Use Case 2		
Name: **User plays a Portal Hunt mission**		
Step	**System Actor Action**	**System Response**
1	Creator access mymultimediaworld game creation page	The system asks for identification
2	Creator authenticates	The system access AT page
3	Creator access Portal Hunt authoring tool	The system give access to the different components of the mission
4	Creator set the teams of the mission	System gives up to 3 teams per session
5	Creator enables mission to be played 1. Stops the mission (if not stopped) 2. Reset the mission 3. Starts mission	System stop, reset the portal and start new mission
6	Gamer access mobile phone GPAC application	System ask for authentication
7	Gamer authenticates himself	The system connects to the server and confirm authentication Success: The system give access to AR games Failure: The system ask to restart
8	Gamer selects a session	System go to the session page
9	Gamer join a team	System starts the game
10	Gamer stats playing: Gamer move around to catch spheres: - Green: Catch One sphere - Red: Catch All sphere - Yellow: Trap sphere  Select Status tab for individual achievement Select Leader tab for leader board (team achievement) Select Camera tab for AR view	System communicates constantly with the server to give the information about the game and the score.



Playing PH game use case diagram

5.7.3 ARQuiz

5.7.3.1 Overview

The ARQuiz is an augmented reality game for mobile devices, fully implemented by using MPEG technologies. Like any quiz, the game consists in answering questions over a given period of time. Unlike a traditional quiz, in order to solve ARQuiz, one has to search for the answers in the physical space. Just look around and, through the camera view of your phone, you'll see virtual hints embedded in the reality. Go next to one hint, collect it and read the story; the answer is between the lines...

5.7.3.2 Example of Play

Bianca, Daniel and Bogdan are attending the 100th MPEG meeting. While doing their registration, they get a flyer about the ARQuiz, an augmented reality game based on MPEG technologies. As they want to test their knowledge about the MPEG life they think it is an excellent idea to take the MPEG100 ARQuiz. They just have to download the ARAF player and game into their Android mobile phones to get started. The first question appears… Bianca does not know the answer and tries to see what Bogdan is answering but they don't have the same

question! Bianca tries then to discover the answer by looking for a hint (under a story format). She takes a look at the map view where the stories are located and starts walking towards one of them. Once there, on her screen a story appears. But it's not helping her; it's not related to the question. She switches then to camera view and now she can see floating numbers embedded in the real word. She finds the right number and heads towards it. Once she is close enough to the number the story pops up magically into her screen… Bingo! Bianca can now answer the question and continue the quiz. She got a total of 40 points. Daniel was not so lucky; he could not finish the quiz on time as he never found the story of the third question. He was too busy following the rabbit that appeared trough the camera view….

Figure 5. Playing the MPEG100 ARQuiz

5.7.3.3 Detailed Description

In ARQuiz, each question has an associated story. In order to get more information about the subject, the players must explore the surroundings using their mobile devices. A map overview shows the game area and the location of all the hints. In AR view mode, the hint positions are illustrated as 3D models embedded in the real world. Once close enough to the hint location, a story will pop up.

After reading the story, the user can display again the question and answer it or he can look for another story.

Figure 6: Map view showing the location of stories (left); 3D models of the stories when viewing in AR (right).

5.7.3.4 Authoring the game

ARQuiz is a novel approach in addressing augmented and mixed reality applications. It is entirely based on a formalism, called ARAF - Augmented Reality Application Format - currently under development in MPEG, that is combining concepts and implementations from MPEG-4 and MPEG-V to create a dynamic, and interactive engaging learning experience. Therefore, the game becomes content and authoring it is as easy as creating content.

A web-based authoring interface allows anyone to create his own ARQuiz. Just select a location where your game should take place and input the number of questions you want to create. For each question, insert the possible answers (text or images), the location of the hint and its story. It is also possible to populate the real world with 3D graphics objects, visible in AR mode.

5.7.3.5 The detailed use case

5.7.3.5.1 *Create an AR Quiz*

Actors

Actor	**Description**
MyMultimediaWorld Authoring tool interface	The authoring tool must be available via the website mymultimediaworld
Browser	Interface used by the game mission creator to create the mission
Game mission creator	Person (can be the gamer or not) that will create the mission, that is the play instance.

Initiator

Game mission creator

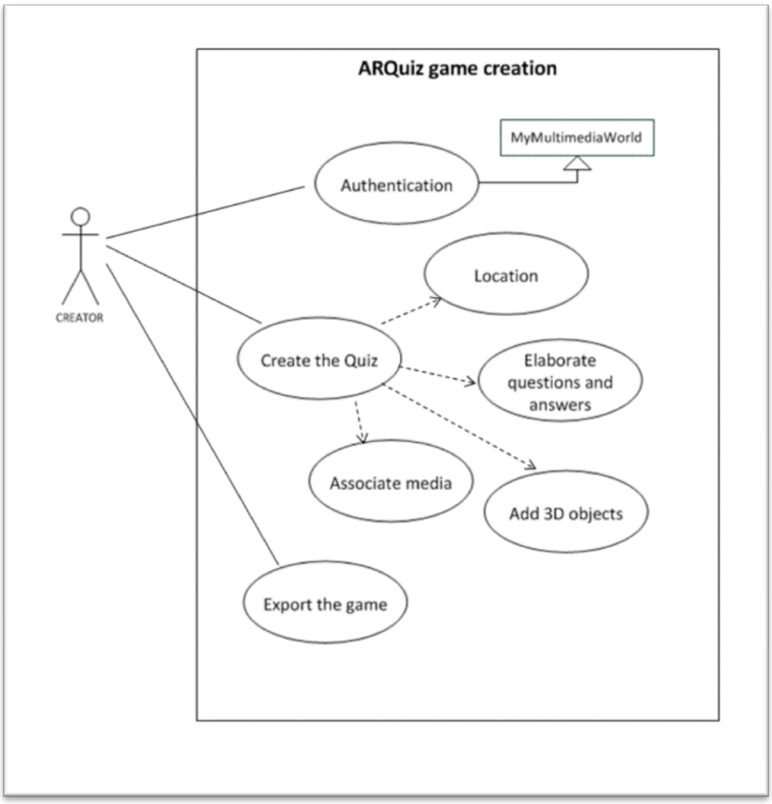
Preconditions

- ARQuiz Authoring Tool is available
- Authentications is made in MyMultimediaWorld.com

Main Flow

Use Case 5		
Name: **User creates a ARQuiz game**		
Step	**System Actor Action**	**System Response**
1	Creator access mymultimediaworld game creation page	System asks for identification
2	Creator authenticates itself	System goes to Authoring tool main page
3	Creator access ARQuiz authoring tool	System displays current quizs and offers possibility of new quiz
4a	Creators creates a new quiz giving the name, the description and the **instructions**	System allows the edition
4b	Creator access existing quiz	System display the mission page with its characteristics
5	Creator edits quiz	System allows the edition
6	Creator determines the location of the question, input the text, the images and the associated answers (specifying which is the right answer) and the question story (hint) There is no limitation in the number of questions	Systems saves the changes and proposes 3D objects as part of the background

7	Creator exports the game	System creates the MPEG file



ARQ game creation use case diagram

5.7.3.5.2 *Playing the ARQuiz Game*

Actors

Actor	Description
MyMultimediaWorld Authoring tool interface	The authoring tool must be available and working for the
Browser	Interface used by the game mission creator to create the mission
Gamer	Person playing the game

Initiator
Gamer

Preconditions

- Authoring Tool is available
- Authentications is made in MyMultimediaWorld.com

- Gamer has the ARAF player installed in his Smartphone
- The Smartphone has GPS integrated

Main Flow

Use Case 6		
Name: **User plays an ARQuiz game**		
Step	**System Actor Action**	**System Response**
1	Gamer download the game	System launches the ARAF player
2	Gamer enters his name	System displays the question
3a	Gamer knows the answer	No interaction with the system
3b	Gamer does not the answer, in order to get more information about the subject, the players must explore the surroundings using their mobile devices	A map overview shows the game area and the location of all the hints. The hint positions are illustrated as 3D models floating in the real world. Once close enough to the hint location, a story will pop-up.
4	Gamer answers the question	A noise will be produced indicating if the answer was wrong or right System displays following question
5	Gamer finishes the quiz	System gives the score (computation of the time and number of correct answers)

Playing ARQ game use case diagram

5.8 *Augmented reality for remote laboratories*

Augmented Reality and, in general, Mixed Reality combines the real world with a virtual scene and creates an immersive and interactive real-time environment. The possible applications are almost limitless and encompass areas like learning, architecture or industrial production. Over the last decade, part of the researches conducted in the field of Online Engineering aimed at controlling real remote appliances over the Internet. Such researches were supported by:

- The industry, for computer-supported maintenance and remote control of processes in the distributed enterprise
- Research institutes, to share expensive devices, mandatory for specific research, yet under-used in term of frequency over the year,
- Distance learning curriculum.

In distance learning, lectures, exercises or projects are easily brought online. It is more difficult with practical laboratories. Meantime, they are a compulsory learning material in engineering studies, since they give students practical skills and know-how. Bringing laboratories online aims at offering learners the possibility to manipulate appliances from different locations. It allows reducing geographical laboratories constraints: ”If you cannot come to the lab, the lab will come to you” [1].

Figure 1 shows a spectrum analyzer. It is a measurement device dedicated to display different frequencies contained in a signal with their respective magnitudes. Signal can be from different natures: electrical, optical, sound or radio-electrical. In engineering studies, such device is unavoidable.

Figure 20, A Spectrum Analyser

Christine Perey 7/22/12 11:53 AM
Deleted: 1

At the Pervasive Computing area, Remote Laboratories are expected to be accessible anytime and anywhere. This is enforced in distance curriculum with different time zones for students, tutors and instrument locations. With the increase of mobile devices computing capabilities, remote devices should be controlled with any kind of terminals able to display multimedia content.

A lot of frameworks for building Remote Laboratories were proposed during the last decade in the literature. Most of them rely on the same architecture: the device to control, a local

machine connected to the device, a middleware and a client application for interacting remotely [2]. Web forms are usually used to post experiment parameters.

On the device side, some standardization efforts already exist. The IVI standards[1] define instrument classes such as Oscilloscope, Spectrum analyzer, Switch, etc., in order to provide interchangeability between instruments of different constructors. The hardware functions, once designed for a class of instrument, are fully compatible with other instruments of the same class, assuming they are IVI compliant. It is not the case for the client side.

Online laboratories designers need to represent feedbacks from the controlled device, after passing commands for instance. Two approaches are generally employed for commands feedbacks. The first one consists in representing the feedbacks from the device, such as graphs, charts, comments etc., as a software component in the Web form. This approach gives the advantage of easily representing the device's feedbacks. The second approach consists in capturing the device itself with a camera and presenting the feedbacks as a natural video in the Web form. The first approach seriously lacks faithfulness of the device's Human Machine Interface (HMI), and the second one entails deploying several technologies to ensure content management (images/videos/graphics/scripts/metadata). Moreover, the hard taste of HCI construction is a story deadlock for remote laboratories development due to resources costs.

Augmented Reality is a view of a real scene where computed information is added. Added Information can be textual, sound, graphics, etc. It can improve the representation's fidelity of Remote Laboratories. Seeing the real scene and interacting with it, i.e. the remote device, leads in having a better comprehension of the manipulated device and improves the "sense of being there", which is an utmost issue in distant learning [3].

5.8.1 Reference

1. V. Judson Harward, Jesus A. Del Alamo, Steven R. Mernam, Philip H. Bailey, Joel Carpenter, Kim- berley DeLong, Chris Felknor, James Hardison, Bryant Harrison, Imad Jabbour, Philip D. Long, Tingt- ing Mao, Loai Naamani, Jedidiah Northbridge, Mark Schulz, Daniel Talavera, Charuleka Varadharajan, Shaomin Wang, Karim Yehia, Rabih Zbib, David Zych, The iLab Architecture: A Web Services Infrastructure to Build Communities of Internet Accessible Laboratories, Proceedings of the IEEE. 96 (2008), 931-950

2. Gravier, C., Fayolle, J., Bayard, B., Ates, M., and Lardon, J. 2008. State of the art about remote laboratories: paradigms- foundation of ongoing mutations. International Journal of Online Engineering (iJOE). 4, 1 (Jan. 2008), 19-25.

3. Tzafestas, C., S. 2006. Virtual and Mixed Reality in Telerobotics: A Survey. *Industrial Robotics: Programming, Simulation and Applications*, Low Kin Huat (Ed.) (Dec. 2006), 437-470.

[1] http://www.ivifoundation.org